

## **REMARKS/ARGUMENTS**

**1.** The amendments made in this reply to the Office communication of June 21, 2005 are intended to:

- (a) cancel new matter, as requested in Item 7 of the Office Action, namely, the new matter identified in Items 5 and 6 of the Office Action;
- (b) revise the claims of the application, in response to Items 11, 15, 16, and 17 of the Office Action; (The above new claims are submitted to be patentable over the art of record for the reasons given below.)
- (c) cancel non-elected claims 2-6 drawn to an invention non-elected, October 20, 2003, in response to Item 21 of the Office Action, as indicated in the instant Listing of Claims.

**2. In response to Item 5 of the Office communication** of June 21, 2005, which is directed to the amendment that Applicant filed on October 20, 2003:

- (a) The item identified as Item 7 in the Office communication is cancelled in Item 2 in the instant Amendment to the Specifications. Item 3 in the instant Amendment to the Specifications is an amendment that crafts a transition sentence using only the originally disclosed matter that had been contained in Item 2.
- (b) The item identified as Item 17 in the Office communication is cancelled in Item 4 in the instant Amendment to the Specifications.
- (c) The item identified as Item 18 in the Office communication is cancelled in Item 5 in the instant Amendment to the Specifications. Item 6 in the instant Amendment to the Specifications is an amendment that deletes the originally disclosed matter that had been contained in Items 4 and 5.
- (d) The item identified as Item 19 in the Office communication is cancelled in Item 7 in the instant Amendment to the Specifications. Item 8 in the instant Amendment to the Specifications is an amendment that crafts a transition sentence using only the originally disclosed matter that had been contained in Item 7.
- (e) The item identified as Item 20 in the Office communication is cancelled in Item 9 in the instant Amendment to the Specifications. Item 10 in the instant Amendment to the Specifications is an amendment that deletes the originally disclosed matter that had been contained in Item 9.

(f) The item identified as Item 24 in the Office communication is cancelled in Item 11 in the instant Amendment to the Specifications.

(g) The item identified as Item 32 in the Office communication is cancelled in Item 12 in the instant Amendment to the Specifications. Item 13 in the instant Amendment to the Specifications is an amendment that crafts a transition sentence using only the originally disclosed matter that had been contained in Item 12.

**3. In response to Item 6 of the Office communication of June 21, 2005, which is directed to the amendment that Applicant filed on July 07, 2004:**

(a) The item identified as Item 17 in the Office communication is cancelled in Item 16 in the instant Amendment to the Specifications.

(b) The item identified as Item 18 in the Office communication is cancelled in Item 17 in the instant Amendment to the Specifications.

(c) The item identified as Item 19 in the Office communication is cancelled in Item 18 in the instant Amendment to the Specifications.

**4. In response to the requirement of Item 7 of the Office communication of June 21, 2005, Applicant has cancelled the identified new matter, in the this reply to the Office Action, as described in the previous two paragraphs.**

**5. In response to Item 11 of the Office communication of June 21, 2005, which describes claim limitations that the Examiner did not find in the instant specification, Applicant responds as follows:**

(a) The Examiner did not find in the specification the limitation "simultaneously ... that also distinguish each subset from the other subset" (lines 1-3), and notes the section "Clustering of Microarray Data" in the instant specification. Applicant responds to the request to identify the written description basis of the limitation as follows.

Page 19, lines 6-13 define gene accession numbers as identifiers used to label the DNA sequences of different genes. Page 27, lines 1-4 defines a clustering of a set of gene accession numbers as an organization of those accession numbers into separate groups. Page 27, lines 10-12 of that section describes more specifically that the data comprising a clustering of gene accession numbers consists of (1) the number of clusters (Cmax) and (2) the accession numbers contained in

each of the Cmax clusters. The disclosure contemplates two situations – either the clustering of accession numbers has been performed by some method that is external to the disclosed invention, or the clustering of accession numbers is derived by the invention's software itself. In the former case (Page 27, lines 5-17), the user of the invention's software provides the clustering data through the contents of a computer file. In the latter case, (Page 27, lines 18-23 and page 28, lines 1-14), the input data to the clustering algorithm are an undivided set of accession numbers, along with gene expression levels corresponding to each of the accession numbers. It is well understood by practitioners of the art that a clustering algorithm puts objects into disjoint groups on the basis of how similar each object is to other members of the group into which the algorithm assigns them (Page 5, lines 7-21). The default algorithm used to derive the clustering data, within the invention's program, is the k-means algorithm, implemented using publicly-available CLARA open source code (Page 29, lines 4-19). Thus, the specification pertains to **a group of subsets of genes** because the set of accession numbers is divided into clusters (each cluster containing a subset of the set of all accession numbers in the undivided list), and there are Cmax such clusters in the group of subsets. More narrowly, the specification could also be said to pertain to a group of subsets of accession numbers, because it is the accession numbers that are clustered. However, because the clustering is performed on the basis of experimental gene expression levels, the accession numbers correspond to genes that have been demonstrated experimentally to be expressed, otherwise they would not included in the undivided set of accession numbers. Hence, Applicant uses within the limitation the term “a group of subsets of genes” rather than “a group of subsets of accession numbers.”

The term “**annotating**” is used within the limitation to mean generating a list of words and phrases that describe something, in this case, that describe each cluster (Page 9, line 20). Thus, the disclosure states that “A further aim of the present system is that it automatically generates a list of words or phrases (‘**annotations**’) that best describe each cluster and that also best distinguishes each cluster from the others.” (Page 9, lines 19-21). In view of the rationale for Applicant's use of the phrase “a group of subsets of genes” presented in the previous paragraph, this aim of the invention's system therefore pertains to **annotating a group of subsets of genes with words and phrases that characterize each subset and that also distinguish each subset from the other subsets**. To that end, Applicant first disclosed methods for obtaining text from

the scientific literature that is associated with each of the clusters. This involves obtaining documents from the scientific literature that are associated with each accession number in the undivided list of accession numbers (Page 21, line 7 through Page 26, line 19). It then involves the grouping of the documents, and their associated text, so as to place sets of sets of the documents, and their associated text, in correspondence with the clusters (Page 36, lines 17-22 and Page 37, lines 1-3). It also involves the creation of a statistical model of the text for all of the text associated with each of the clusters **simultaneously**, by using a single system function call to instruct the publicly available, open source code computer program 'rainbow' to do so (Page 37, lines 3-8). Finally, it involves a system function call to the computer program 'rainbow' to obtain the word/phrase annotation lists for each of the clusters, making use of rainbow's implementation (McCALLUM, 1998) of classification, log-odds ratio and mutual information algorithms that are intended to distinguish each class, defined by a cluster, from one another in terms of the word/phrase annotation lists for each cluster (Page 38, line 4-22 and Page 39, lines 1-15).

The Examiner's concern may be that the disclosure involves of a sequence of steps, not a single step in which all operations occur simultaneously. Applicant meant in the recited limitation that there is simultaneity in the derivation of word/phrase annotations for the entire group of subsets of genes, by virtue of facts that the computer program 'rainbow' creates its statistical model of the text for all of the clusters through a single system call to 'rainbow', and by virtue of the fact that algorithms such as mutual information involve calculations for each cluster that take into account information pertaining to all of the clusters at once. Because the term "simultaneously" is ambiguous in this context, that wording of lines 1-3 is now amended in the current Listing of Claims by omitting the word "simultaneously". The wording of the claim is also amended by using the phrase "subset of genes" rather than simply "subset" in response to Item 15 of the Office Action, and in response to Item 17 of the Office action because the claim is also concerned with another type of set, namely, a set of literature documents.

**(b)** The Examiner did not find in the instant specification the limitation "of genes that generally have not sequence similarities among them" (line 4). Applicant responds by noting that wording of step (a) in Claim 1 is now amended in the current Listing of Claims to read simply "means for identifying a set of genes". The specification does not restrict the types of genes that may be

found in the set of genes that are identified, so there is no need to indicate any type of limitation related to sequence similarities among genes in the set. On page 21, line 3-4, the specification simply describes, as input to the invented computer program, any computer file containing a set of gene accession numbers corresponding to the DNA species that have been spotted onto a microarray.

(c) The Examiner did not find in the instant specification the limitation of “the computer program Rainbow that are intended to simultaneously ... to distinguish each cluster from the others” (lines 14-15). The Examiner was correct in objecting because in the previous Listing of Claims, the antecedent of “that” in the phrase “that are intended” is vague. The Examiner appears to have read the previous wording of the claim to mean “methods that are implemented in the computer program Rainbow that are intended...”, whereas Applicant meant “assignment ... followed by the application ..., that are intended ...”. Wording of step (e) in Claim 1 is now amended in the current Listing of Claims in response to Items 15 and 17 of the Office Action.

**6. In Item 15 of the Office communication of June 21, 2005, the Examiner finds that the limitation of “distinguish each subset from the other subsets”(line 3), “the others”(line 16), or “the other clusters” (lines 24-25), respectively, causes claim 1 to be vague and indefinite because it is not clear as to what suitability characteristic is being evaluated to distinguish a “characterized each subset” from the “other subsets”, “the others”, or “the other clusters”. Applicant responds by noting that the wording throughout claim 1 is amended in the current Listing of Claims to use the phrase “subset of genes” throughout the claims rather than treat that phrase as being interchangeable with the term “cluster”. When one “subset of genes” is being compared with another “subset of genes”, the revised limitation now makes clear that an individual subset of genes is being distinguished from other individually considered subsets of genes in terms of different words and phrases that the system attaches to different individual subsets of genes.**

**7. In Item 16 of the Office communication of June 21, 2005,** the Examiner finds that clarification of metes and bounds is required, in connection with the limitation of line 4 in which genes “generally have no sequence similarities among them”. Applicant responds by noting that wording of step (a) in Claim 1 is now amended in the current Listing of Claims to read simply “means for identifying a set of genes”. The specification does not restrict the types of genes that may be found in the set of genes that are identified, so there is no need to indicate any type of limitation related to sequence similarities among genes in the set. On page 21, line 3-4, the specification simply describes, as input to the invented computer program, a computer file containing a set of gene accession numbers corresponding to the DNA species that have been spotted onto a microarray.

**8. In Item 17 of the Office communication of June 21, 2005,** the Examiner finds that steps (c) and (d) of Claim 1 are vague and indefinite because it is not clear whether step (c) is directed to two sets of documents or one, and the antecedent basis for the limitation of “the documents” in line 11 of step (e) is not clear. Applicant agrees that the claims as written were not clear. Accordingly, wording of steps (c), (d), and (e) in Claim 1 are now amended in the current Listing of Claims to make it clear that “the documents” that had been referred to in line 11 were “a set of documents” that had been referred to in line 7, consisting of a superset of the “set of literature” documents for each gene that had been referred to in line 6. The antecedent basis for the limitation of “the documents” that had been referred to in line 11 of step (e) is now made clear in the current Listing of Claims, by moving a rephrased description of the limitation from the earlier step (e) into step (d).

**9.** Applicant requests pursuant to MPEP 707.07(j) that the Examiner draft one or more suitable claims for the applicant, if the Examiner finds patentable subject matter disclosed in this application, but feels that Applicant’s present claims are not entirely suitable.

10. Applicant thanks the Examiner for thoughtful consideration of this application as expressed in the Office action, because Applicant believes that his responses to the Examiner's action would strengthen the application.

Respectfully submitted,

David R Rigney

David R. Rigney, Inventor

GENETWORKS Inc.

P.O. Box 33296

Austin TX 78764-0296

Tel. 512-445-7301

drigney@genetworks.com